18. SEPARATION STATISTICS

The person and item separation statistics used in Rasch measurement are still unfamiliar to some practitioners (Wright, Mead and Bell, 1979; Wright and Stone, 1979; and Wright and Masters, 1982). This primer describes the separation statistics and illustrates some of the key concepts and procedures of separation by working through an example.

A variable can be thought of as a straight line. To measure successfully we must be able to locate both items and persons along this line. Items are located by the number of persons getting a specific item correct. Persons are located by how many items they were able to answer correctly. Items to the left on the line are easier than those to the right while persons to the left have less ability than others to the right.

It is necessary to locate persons and items along the variable line with sufficient precision to "see" between them. Items and persons must be separated along this line for useful measurement to be possible. But, separation that is too wide usually signifies gaps among item difficulties and person abilities. This leads to imprecise measurement. Separation that is too narrow, however, signifies redundancy for test items and not enough differentiation among person abilities to distinguish between them.

Items must be sufficiently well separated in difficulty to identify the direction and meaning of the variable. To be useful, a selection of items, a test, must separate relevant persons by their performance. The item locations are the operational definition of the variable of interest while the person locations are the application of the variable to measurement.

The item and person separation statistics in Rasch measurement provide an analytical tool by which to evaluate the successful development of a variable and with which to monitor its continuing utility. Person separation indicates how efficiently a set of items is able to separate those persons measured. Item separation indicates how well a sample of people is able to separate those items used in the test. Where these statistics are expressed as reliabilities, they range from 0.0 to 1.0. The higher the value the better the separation that exists and the more precise the measurement.

PERSON SEPARATION RELIABILITY

The data in Table 18.1 are from the calibration output of the BICAL program developed by Wright and Mead in 1976. This example is based upon the calibration of 14 Knox Cube tapping items taken by a sample of 34 persons. These data are also discussed in Wright, Mead and Bell, 1976 and Wright and Stone, 1979. At the bottom of the table the Person Separation Reliability (PSR) is given as 0.68. This indicates that the 14 items used in this version of the Knox Cube Test were able to separate the 34 people tested to a moderate degree.

The PSR reported by most Rasch computer programs is calculated by subtracting the ratio of the sample mean square person measure error (MSEp) to the sample person measure variance (SD_p^2) from one. The formula (Wright and Masters, 1982, p.106) is:

Figure 18.1

Original responses of 35 persons to 18 items on the Knox Cube Test.

	Item Name																		
PERSON NAME	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	PERSON SCORE
1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	7
2	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	10
3	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	10
4	1	1	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	6
5	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	10
6	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	10
7	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	14
8	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	10
9	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	10
10	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	11
11	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	8
12	1	1	1	1	1	0	1	0	1	1	0	0	0	0	0	0	0	0	8
13	1	1	1	1	1	0	0	1	1	1	1	1	0	0	0	0	0	0	10
14	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	11
15	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	13
16	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	10
17	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	9
18	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	11
19	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	9
20	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	11
21	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	12
22	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	12
23	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0	12
24	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	0	14
25	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	5
26	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	10
27	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	7
28	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	10
29	1	1	1	1	1	1	0	0	1	1	1	0	0	1	0	0	0	0	10
30	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	9
31	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	10
32	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	11
33	1	1	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	6
34	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	0	0	0	12
35	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
ITEM SCORE	35	35	35	32	31	30	31	27	30	24	12	6	7	3	1	1	1	0	

Wright & Stone, Best Test Design (Chicago: MESA Press, 1979), p.31.

$$PSR = 1 - \left[\frac{MSEp}{SDp^2}\right] \qquad 18.1$$

Table 18.1 shows the data necessary to calculate the PSR for this example. Using these data the equation becomes:

$$PSR = 1 - \frac{32.99}{104.93} = 0.68 \qquad 18.2$$

This Person Separation Reliability is comparable to the KR20 measure of internal consistency. The PSR can be corrected for degrees of freedom [*L*/(*L*-1)] and yield a result which is very similar to that of the KR20. With this correction the PSR becomes:

$$PSR = \left[1 - \frac{32.99}{104.93}\right] \times \left[\frac{13}{12}\right] \qquad 18.3$$

$$PSR = 0.74$$

This "corrected" Person Separation Reliability is close in both formulation and value to the KR20 reliability coefficient. In order to compare the PSR to the KR20, we need to return to the original item-by-score matrix analyzed by the Rasch computer program, BICAL. This matrix is given in Figure 18.1 and comes from Wright and Stone, 1979, p.31. Calculation of the KR20 on this 14 x 34 matrix produces a KR20 of 0.72. The difference between this KR20 of 0.72 and the "corrected" PSR of 0.74 is due to the curvilinear relation between the nonlinear raw scores on which the KR20 is based and the linear logit measures on which the PSR is based.

These calculations of the PSR of Rasch psychometrics and the KR20 on the same data matrix illustrate their equivalence and elucidate the calculation of person separation reliability.

The similarity of these results can be seen in the similarity of the formulae:

$$KR20 = \left[\frac{L}{L-1}\right] \times \left[1 - \frac{\sum pq}{\sigma_y^{\ 2}}\right] \qquad 18.4$$

and

$$\text{Rasch } PSR = \left[\frac{L}{L-1}\right] \times \left[1 - \frac{MSEp}{SDp^2}\right] \qquad 18.5$$

In these formulae *pq* is the error variance of the response score of a "person" for whom the sample item p-values apply, while MSE*p* is the sample average measure error variance in logits and $\sigma_y^{\ 2}$ is the sample variance of the nonlinear raw scores, while $SD_p^{\ 2}$ is the sample variance of the linear logit measures.

Table 18.1

Knox Cube Test Output from Bical

Raw Score	Count	Logit Ability	Standard Error
11	2	3.31	0.92
10	1	2.53	0.93
9	4	1.71	0.96
8	5	0.81	1.03
7	12	-0.22	1.07
6	3	-1.19	0.97
5	2	-1.96	0.86
4	2	-2.61	0.81
3	2	-3.21	0.81
2	1	-3.86	0.88

Person Separation Reliability = 0.68 (without D. F. Correction)

(Best Test Design, Page 57, Table 3.4.2)

ITEM SEPARATION RELIABILITY

KR20 is commonly calculated for items, but almost never for persons. When Hoyt (1941) published his paper on test reliability by ANOVA, he recognized both approaches saying "extended examination of the 'among' items variance would make it possible to decide on the heterogeneity of the respective difficulties of the items while a more extended examination of the 'among students' variance would make it possible to answer certain pertinent questions regarding the individual differences among students" (page 156). This good advice, however has never been followed in practice.

In Rasch measurement, however, the Item Separation Reliability (ISR) is routine (Wright and Masters, 1982, p.92). This ISR gives the test user an indication of how well items are separated by the persons taking the test. The formula for this index is:

$$ISR = 1 - \left[\frac{MSE_I}{SD^2{}_I} \right] \qquad 18.6$$

This is calculated in a fashion similar to the Person Separation Reliability. The higher the ISR, the better those particular items are separated by the persons taking the test.

It is not the algebraic and statistical similarity of the KR20 and Rasch PSR, however, that is of major importance now that it has been demonstrated. Instead, it is the decomposition of these single indices into their constituent parts that leads to a more detailed and more useful management of the test characteristic traditionally referred to as "reliability."

With Rasch calibration we are able to obtain the standard error of calibration for each individual item as well as the standard error of measurement for each person ability. With traditional methods, a standard error of measurement is provided only for measures at the group mean of person ability.

The standard error specific to each item (or person) statistic is far more useful than any single sample (or test) "average". The location of each item and person on a line representing the variable together with their standard errors shows us the definition and utility of the variable. The definition of the variable is specified by the location of the items. The utility of the variable for measuring persons is quantified by the standard error which accompanies each person measure.

MEASUREMENT ESSENTIALS

2nd Edition

BENJAMIN WRIGHT

MARK STONE

WIDE RANGE, INC.
Wilmington, Delaware